Complex genetic regulation of protein glycosylation

Complex genetic regulation of protein glycosylation

Gordan Lauc,*[ab] Igor Rudan,[cde] Harry Campbell[ef] and Pauline M. Rudd[g]

Received 29th May 2009, Accepted 21st September 2009
First published as an Advance Article on the web 19th October 2009
DOI: 10.1039/b910377e

One hundred years have passed since Archibald Garrod postulated the one gene/one enzyme hypothesis. Since then, science has made significant progress and geneticists are now tackling an overwhelming complexity of gene regulation networks that underlie the genetics of complex human diseases. A particularly complex element in the biology of higher organisms is the genetics of protein glycosylation. Nearly all proteins that appeared after the emergence of multicellular life are glycosylated, but instead of being molded by a single gene, glycan structures are encoded within a network of several hundred glycosyltransferases, glycosidases, transporters, transcription factors and other proteins. In addition, in contrast to the linear structures of DNA and proteins, glycans have multiple branches that make their analysis significantly more challenging. However, recent developments in high throughput HPLC analysis have advanced glycan analysis significantly and it is now possible to address questions about the complex genetics of protein glycosylation. In this review we present some preliminary insights into this fascinating field.

[a] *University of Zagreb, Faculty of Pharmacy and Biochemistry, A. Kovacica 1, HR-10000 Zagreb, Croatia. E-mail: glauc@pharma.hr; Fax: +385 1 639 4400*
[b] *Genos Ltd., Glycobiology Division, Planinska 1, Zagreb, Croatia*
[c] *Gen Info Ltd., Zagreb, Croatia*
[d] *Croatian Centre for Global Health, University of Split Medical School, Split Croatia*
[e] *Department of Public Health Sciences, The University of Edinburgh Medical School, UK*
[f] *Institute of Genetics and Molecular Medicine, The University of Edinburgh, UK*
[g] *Dublin-Oxford Glycobiology Lab., NIBRT, Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland*

1. Structural and functional aspects of protein glycosylation

Carbohydrates have been a focus of human interest since the very beginning of civilization, but we have only recently started to understand the importance of complex oligosaccharides (glycans) attached to protein or lipid backbones. This is perhaps not surprising, since the branched structures of sugars make the analysis of glycoconjugates significantly more challenging than the analysis of linear DNA and protein sequences.

Glycosylation is the only post-translational modification that can produce significant structural changes to proteins.

Gordan Lauc

Gordan Lauc is Professor of Biochemistry and Molecular Biology at the University of Zagreb Faculty of Pharmacy and Biochemistry and University of Osijek School of Medicine. He graduated Molecular Biology at the University of Zagreb in 1992, and obtained PhD in Biochemistry from the same university in 1995. After visiting Johns Hopkins University in 1996/97 as a Fulbright Fellow he was appointed visiting professor in the JHU Biology Department between 2001 and 2007. Gordan Lauc is author of over 70 research papers published in international journals and co-inventor on five international patents. His main research interest is the role of glycans and their interaction with protein receptors in the pathophysiology of different diseases associated with stress.

Igor Rudan

Igor Rudan graduated from the University of Zagreb Medical School, Croatia, in 1995. He obtained an MSc in Anthropology (1997) and a PhD in Epidemiology (1998) from Zagreb University; M.P.H. from the University of Pavia, Italy (1999); and a 2nd PhD in genetics at the University of Edinburgh, UK (2005). He is affiliated to the University of Edinburgh, UK and is also a Professor and Director of Croatian Centre for Global Health. He published more than 100 papers and 7 books/monographs to date, and has been awarded 15 national and international awards. He has been (co)principal investigator on 20 research grants. He is as a consultant of the World Health Organization, UNICEF, DFID (UK) and Child Health and Nutrition Research Initiative (CHNRI).

A typical glycan is a complex molecule containing between 10 and 15 monosaccharides linked in a rather complicated manner. Between two and five glycans are attached to an average glycoprotein, resulting in an exceedingly complex glycoproteome, estimated to be several orders of magnitude more complex than the proteome itself.[1]

In eukaryotes, glycosylation occurs through 11 biosynthetic pathways.[2] The oldest pathway, which is ubiquitously present, is N-glycosylation. This occurs when a block of 14 sugars (the dolichol phosphate precursor) is transferred co-translationally to specific asparagine (Asn) residues in newly synthesized polypeptides in the endoplasmic reticulum.[3] The resulting N-glycans are subjected to extensive modification as glycoproteins mature and move *via* the Golgi complex to their intra- and extra-cellular destinations.[4]

Depending on the structure of a glycan and the protein to which it is attached, glycans can have many different functions: they can be important in the proper folding of proteins; they can regulate the function of protein backbones by differential processing of glycosylation;[5] they can be strategically placed so that they are able to provide protease protection without interfering with the function of the protein; they can serve as recognition motifs for specific carbohydrate binding proteins (lectins) and mediate cell–cell interactions;[6] they can enable proteins and lipids to "jump" from one cell to another;[7] and they can also have many other functions, some of which are still poorly understood.[8]

The complex shape, functionality and dynamic properties of glycans allow these molecules to function in intermolecular interactions as encoders of biological information. Carbohydrate recognition is an integral part of normal biological development[9] and the immune defence against pathogens is mediated through the identification of exogenous carbohydrates.[10] Cells of higher eukaryotes are covered with a dense layer of glycoconjugates (glycocalyx) which are responsible for their communication with the outside world. Many bacterial and viral pathogens initially adhere to host tissues by binding specifically to carbohydrates on the host cell surfaces,[11] thus the heterogeneity of glycan structures represents a valuable tool that higher eukaryotes use to outmanoeuvre rapidly evolving pathogens. Evolutionary processes have employed the potential for variability offered by glycosylation and have resulted in the enormous complexity in glycan structures that we observe today in multicellular organisms. The transition from a single cell to multicellular organisms required the invention of many new molecules and mechanisms, needed to guide and control interactions between individual cells and their surroundings including other cells, the extracellular matrix, and pathogens. Glycosylation was apparently a very useful tool in this evolutionary process, and today, nearly all membrane and extracellular proteins are glycosylated.[12]

Even though the mass content of glycans in a glycoprotein is often rather small (5–10%), since glycans have a much lower density than proteins, glycans always represent a significant part of the hydrodynamic volume of the glycoprotein. In spite of differences in genealogy and biosynthetic mechanisms of the carbohydrate and protein entities that constitute a glycoprotein, once synthesized, the glycoprotein functions as a single unit and other interacting molecules cannot differentiate whether they are binding to a protein, or a carbohydrate part of the molecule.[13] Structural and conformational aspects of glycans are very complex,[14] and a small change in a glycan structure can have important functional consequences. Therefore, in order to comprehend the functions of glycoconjugates, it is necessary to determine the structure and function of their glycans. For example, it is difficult to imagine that the function of the prion protein (Fig. 1) could be fully explained without understanding its glycan part.

N-glycosylation is essential for multicellular life and its complete absence is embryonically lethal.[15] Genetic defects

Harry Campbell

Harry Campbell is Professor of Genetic Epidemiology and Public Health, and Head of Public Health Sciences, University of Edinburgh. His primary research interests are in colorectal cancer genetics and genetics of complex disease in genetic isolate populations. He has published over 200 original articles, and numerous editorials and book chapters. He is PI and co-ordinator of the European research project funded through FP6 called EUROpean Special Populations reseArch Network (EUROSPAN): quantifying and harnessing genetic variation for gene discovery. This is carrying out QTL mapping of a wide range of traits including psychological and cognitive traits directly relevant to this application in several European populations. He has recently completed a GWAS in colorectal cancer which identified several novel common variants influencing risk of this condition.

Pauline M. Rudd

Pauline Rudd is NIBRT Professor of Glycobiology at University College, Dublin. She obtained a BSc in Chemistry (London) and a PhD in Glycobiology (Open University) and, before bringing her group to Dublin, was a University Reader in Glycobiology at Oxford University. She was a Founding Scientist of Wessex Biochemicals (later Sigma London), Visiting Research Associate at The Scripps Research Institute, CA, Visiting Professor at Shanghai Medical University PRC, Visiting Scientist at Ben Gurion University, Israel, Erskine Visiting Fellow, Canterbury University, NZ, is a fellow of the RSM and adjunct Professor at North Eastern University, Boston.

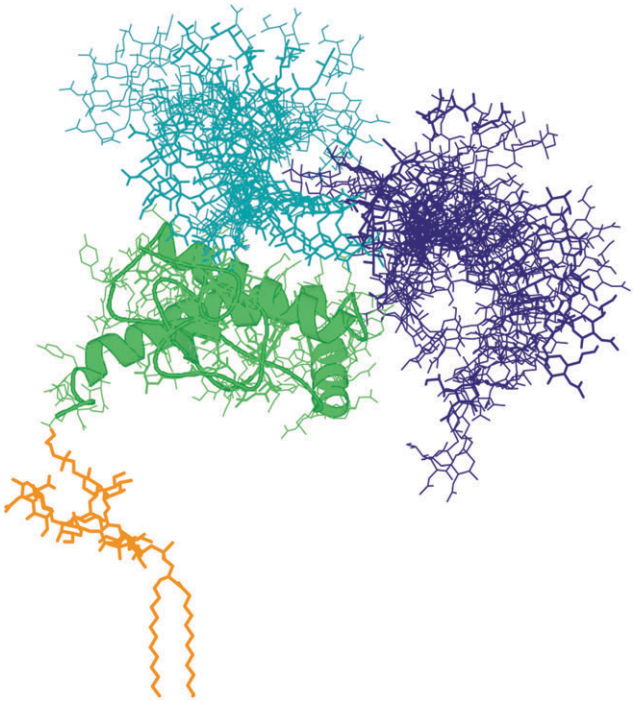

Fig. 1 Prion protein (PrP^C). Prion protein is a glycoprotein composed of the amino acid backbone (green), two N-glycans at Asn 181 (light blue) and Asn 197 (dark blue), and a glycosylphosphatidylinositol anchor (orange). The N-glycans are modeled to show the range of conformational space available to the sugars. Figure courtesy of M. R. Wormald, R. A. Dwek and P. M. Rudd, Glycobiology Institute, Oxford, UK.

that affect protein glycosylation are relevant to human health and cause a number of different human diseases.[16,17] In addition, it has been shown that a surprisingly high share of all known disease-causing missense mutations result in gains of potential glycosylation sites.[18,19] Contrary to core glycans, which appear to be essential for the functions of many glycoproteins, variability in monosaccharides at the end of glycan antennas is common (*e.g.*, ABO blood groups). This structural diversity of glycans contributes to glycoprotein heterogeneity in a population that can be advantageous for evading pathogens and adapting to changing environment.[8]

One example of a protein where structural and functional aspects of glycosylation have been studied in detail is the immunoglobulin G (IgG) molecule. Glycan structures attached to the Asn_{297} residue in the constant (CH2) domains of IgG are critical for maintaining the overall structure and function of the molecule.[20] For over 20 years, it has been known that a decrease in the number of galactose residues at the non-reducing termini of these glycans is associated with various diseases.[20–22] Under-galactosylation in rheumatoid arthritis apparently precedes the outbreak of other symptoms, and, when clustered, agalactosyl glycoforms may initiate the complementary pathway, thus contributing to the pathogenesis of the disease.[23] IgG mediates both pro- and anti-inflammatory activities through the engagement of its Fc fragment with distinct receptors. These distinct properties of the IgG Fc result from differential glycosylation.[24] IgG acquires anti-inflammatory properties upon Fc sialylation, which is reduced upon the induction of an antigen-specific immune response. This differential sialylation may provide a switch from innate anti-inflammatory activity in the steady state to generating adaptive pro-inflammatory effects upon antigenic challenge.[25] Glycosylation of IgG also plays an important role in mediating antibody-dependent cellular cytotoxicity (ADCC) following treatment with therapeutic antibodies, and this is closely related to the clinical efficacy of anticancer drugs. The lack of core fucose in glycans

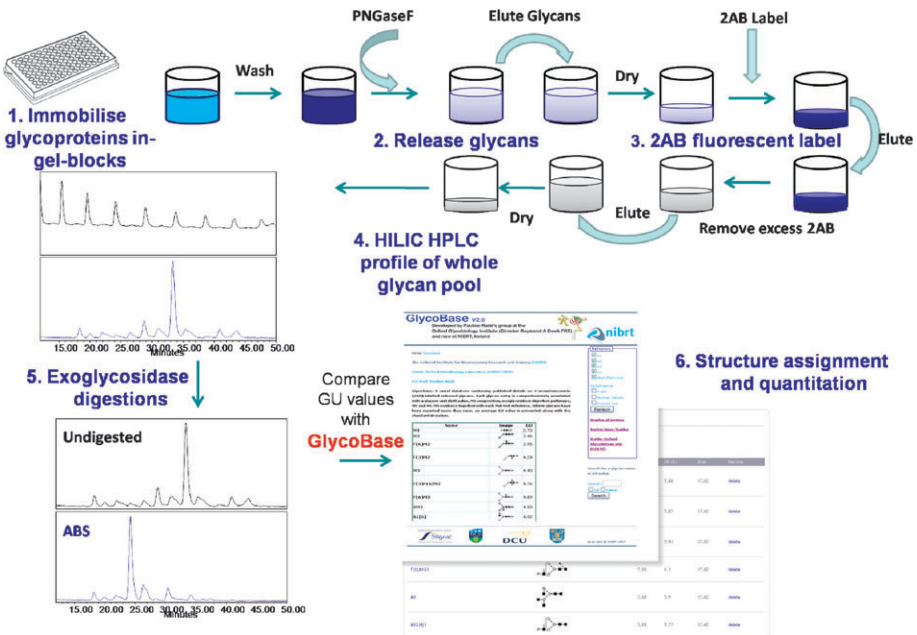

Fig. 2 Platform for robotics-compatible high-throughput HPLC-based method for glycan analysis with computer-assisted data interpretation (Dublin-Oxford Group, NIBRT, Ireland).

attached to the Fc region of antibodies improves Fc gamma RIIIa binding, allowing the antibodies to enhance dramatically the antibody effector functions of ADCC.[26]

2. Variability and heritability of glycans

Glycans display a much higher variability between species than proteins,[27–29] indicating that they evolve much more rapidly. Until recently, due to the technological challenges, detailed quantitative analysis of total plasma glycans was generally performed on a limited number of samples;[30] thus, insights into glycome variability at the population level are very limited. Using a newly developed 96-well plate high-performance liquid chromatography platform (Fig. 2),[31] we have demonstrated that human plasma N-glycans can be separated into chromatographic peaks and quantified reliably in a relatively high-throughput manner.[32] This represents a major technological advance, one which was required to investigate the genetic regulation and biological role of glycan structures and to bring glycomics into line with genomics, proteomics and metabolomics.[33] Other methods that enable high-throughput glycan analysis, like mass spectrometry (MS) or capillary electrophoresis (CE), were until now not shown to be adequate for this type of study. The main problem in MS analysis of glycans is the limited ability to reliably quantify diverse glycan structures, while capillary electrophoresis lacks sufficient resolution to analyze complex mixtures.

By analyzing the distribution of plasma glycans in a sample of 1008 individuals, we observed an unusually large biological variability at the population level, with a median ratio of minimal to maximal values of 6.2. Due to experimental limitations, in that study we did not quantify individual glycan structures, but instead measured glycans partitioned into 33 groups containing several (mostly similar) glycan structures.[32] Despite the fact that different glycans within the same group could have changed in opposite directions, thus diminishing the cumulative change, the variability of glycans in the population was found to be high. This apparently large variability of glycan structures should not be ignored in the development of diagnostic tools based on measuring alterations in glycan levels, since the magnitude of reported changes associated with a disease[30,34–37] is sometimes smaller than the observed natural variability revealed in this study. However, it should be noted that this study only compared ratios of plasma glycans grouped into 33 chromatographic peaks and the possible presence or absence of some less abundant structures in specific individuals was not examined.

Variations observed in a human phenotype are generally due to the joint influence of genetic differences and varying exposures to environmental factors. Heritability is one of the most basic and often one of the first analyses made in a genetic study. Since it represents the proportion of the trait variance that can be attributed to genetic factors, it is often used as a first screening tool to assess whether a trait may be suitable for gene mapping.[38] In our recent study, we performed the first estimation of heritability for a range of different N-glycans. A broad range of variation in heritability was observed, from insignificant or very low to high, indicating that while some glycans are mostly controlled by genetic factors, others are mainly under environmental control. Heritability is important since it can provide information about the underlying genetic structure of the trait, by reporting on additive and dominant variance components.[39,40] However, the interpretation of heritability estimates is complicated by a number of factors.[38] Differences in methodological approaches and sampling schemes, variation in trait values over time, and population-specific differences can all influence heritability values,[38] resulting in a wide range of heritability estimates in different studies. The latter is largely due to the fact that since heritability is (in the simplest form) calculated as the ratio of genetic over genetic and environmental variance, its estimation depends on the environment from which the sample has been collected. Care should therefore be taken in comparing heritability estimates from different populations.

Age, gender and age–gender interaction effects were found to be statistically significant in the majority of the glycan groups, although their effect was relatively small, generally in the range of up to 10%.[32] A number of environmental variables (such as smoking) were also found to be associated with plasma glycome components in a very complex manner. These variables should ideally be measured and included in the analysis of future studies investigating the relationship between plasma glycan levels and disease or disease traits.

3. Advances in genetic investigation of common complex diseases and disease traits

Until about three years ago, genetic linkage studies using scarce sets of genetic markers and candidate-gene approaches based on *a priori* hypotheses on biological plausibility were the major designs used to study the role of human genetic variation in common complex diseases of late onset. However, the linkage approach has only been successful in identifying rare genetic mutations with high penetrance and large effects in extended pedigrees affected by monogenic (“Mendelian”) disorders. These studies were grossly underpowered to detect more common genetic variants of smaller effect sizes in the samples drafted from general population.[41] The candidate gene studies have typically used sample sizes of several hundred cases and controls and the vast majority of reported positive results over the past decade have never been replicated in subsequent studies.[42,43]

More recently, however, a major technological breakthrough occurred. Advances in genotyping technologies were combined with the results of the HapMap project, which identified and validated more than 3 million of the new class of genetic markers—“single nucleotide polymorphysms” (SNPs). The selection (“haplotype tagging”) of several hundred thousand of a particularly useful subset of these SNP markers was then performed, taking into account recombination hotspots in the genome and their linkage disequilibrium (LD) patterns. This has resulted in the development of high density SNP arrays with hundreds of thousands of SNP markers spanning across the entire genome, which made hypothesis-free genome-wide association studies (GWAS) possible.[44]

In parallel with these advances, appropriate statistical and computational methods were developed to deal with issues such as population stratification, relatedness of the subjects in the sample and multiple testing, with agreement on a very stringent genome-wide statistical significance level of approximately $p = 5 \times 10E\text{-}7$ when a set of about 300 000 SNPs is used, to avoid false-positive results that plagued the field for more than a decade.[45,46] These were conducted initially as separate studies in reasonably small sample sizes of up to 2000 cases and controls, but initial successes soon resulted in the formation of very large consortia and pooling of the samples among several studies in genome-wide association study meta-analyses. Current sample sizes of the largest consortia, such as GIANT (which studies anthropometric traits) or DIAGRAM (which studies type 2 diabetes), reach about a hundred thousand cases and controls. Hundreds of common genetic variants underlying complex human diseases and biomedically relevant quantitative traits have been identified, and widely replicated across populations. Genetic networks influencing a wide range of common complex diseases have begun to emerge, including type 1 and 2 diabetes, ischaemic heart disease, stroke, rheumatoid arthritis, breast/prostate/colorectal/lung cancer and inflammatory bowel disease, and a range of endophenotypes underlying these diseases, such as obesity, LDL cholesterol and fasting glucose. These variants typically had small effect sizes (ORs of 1.1–1.4) and many have been located in regions outside the known genes, presumably in distant regulatory regions. These variants individually only explain a very small percentage of the overall variance in disease or traits and thus cannot be used at this point in time for a reliable genetic profiling of disease risk in individual subjects. However, many new aetiological hypotheses have been generated and these GWAS have formed a robust base of new knowledge on which functional studies can now build to explore new disease pathways.[47–49] This should all eventually result in improved prediction of disease risk and a highly personalized era of medicine with new class of drugs that will exploit the new knowledge generated by these studies.

4. Complex genetics of protein glycosylation

Glycosylation is the most diverse post-translational protein modification. Hundreds of specific enzymes and other proteins are involved in the synthesis of complex glycans that are covalently bound to protein backbones.[50] Some of these enzymes are very specific and contribute to the synthesis of a limited number of structures on a small number of proteins,[51] while others (like ER-glycosyltransferases involved in synthesis of N-glycans, or pathways that produce sugar nucleotides) affect thousands of different proteins. This process is not template driven and is thus prone to the influence of various different internal and external factors. The intricate mechanisms by which the interplay of gene expression and intracellular localization of their products give rise to specific glycan structures is only starting to be understood.[52]

Contrary to the polypeptide backbone, glycan parts of glycoproteins are not directly encoded in genes. Instead of being moulded by a single gene, glycan structures are encoded in a network of hundreds of glycosyltransferases, glycosidases, transporters, transcription factors and other proteins. The invention of glycosylation represented a significant evolutionary advantage since it can produce novel structures by modifying gene expression, intracellular localization and activity of biosynthetic enzymes. This invention occurred in parallel with the appearance of multicellular life, and even though it is too speculative to claim that the invention of glycosylation enabled the appearance of multicellular life, the fact that nearly all proteins that came into existence after the appearance of multicellular organisms are glycosylated, indicates the importance of glycosylation for complex organisms.

Due to technological limitations in studying branched chains of glycans, the knowledge about genes and genetic polymorphisms that regulate glycosylation is mostly limited to a small number of glycosyltransferases that directly synthesize glycans.[16] The number of functional polymorphisms in these genes appears to be limited since each change in them affect thousands of proteins and have important consequences for survival.[16] In addition to glycosyltransferases, numerous other proteins like transcription factors, transporters and Golgi organizers are involved in protein glycosylation. The use of these proteins frequently varies between different cell types and physiological states; thus, mutations in their genes would not be as deleterious at the level of the organism. Consequently, it is likely that polymorphisms in these genes might be even more important for determining variations in levels of individual glycan structures in human populations, than polymorphisms in genes for glycosyltransferases.

The recent technological advances in both human genomics and glycomics analysis described above have now made it possible to bring together high throughput genomics and high-throughput glycomics and to apply the successful design of genome-wide association studies to investigate genetic regulation of the human glycome. After demonstrating that glycan measurements were reproducible and that plasma N-glycan levels are heritable,[32] we recently attempted the first genome-wide association study of human plasma N-glycan levels using the Illumina Human Hap300 platform (comprising 317 503 SNP). These studies are still in progress and the results need to be replicated, but they appear to be very successful. As a proof of principle that glycans measured by HPLC are amenable to genome-wide association studies, we recently reported the identification of polymorphisms that affect levels of the A2 glycan (GlcNAc2Man3GlcNAc2) in human plasma proteins.[53] The most significant association was found with a single nucleotide polymorphism rs7161123 located in intron 2 of the gene *fucosyltransferase 8* (Fig. 3). Allele *A* of this SNP alone explained 3.93% of the variance of glycan A2 (after adjustment for age and sex).

5. Conclusions and future prospects

The past few years have witnessed an exponential increase in the level of knowledge about the importance of glycans in numerous physiological and patho-physiological processes. The recent development of new robust high-throughput methods that enable reliable quantitative analysis of numerous samples paved the way for high-throughput glycomics to join

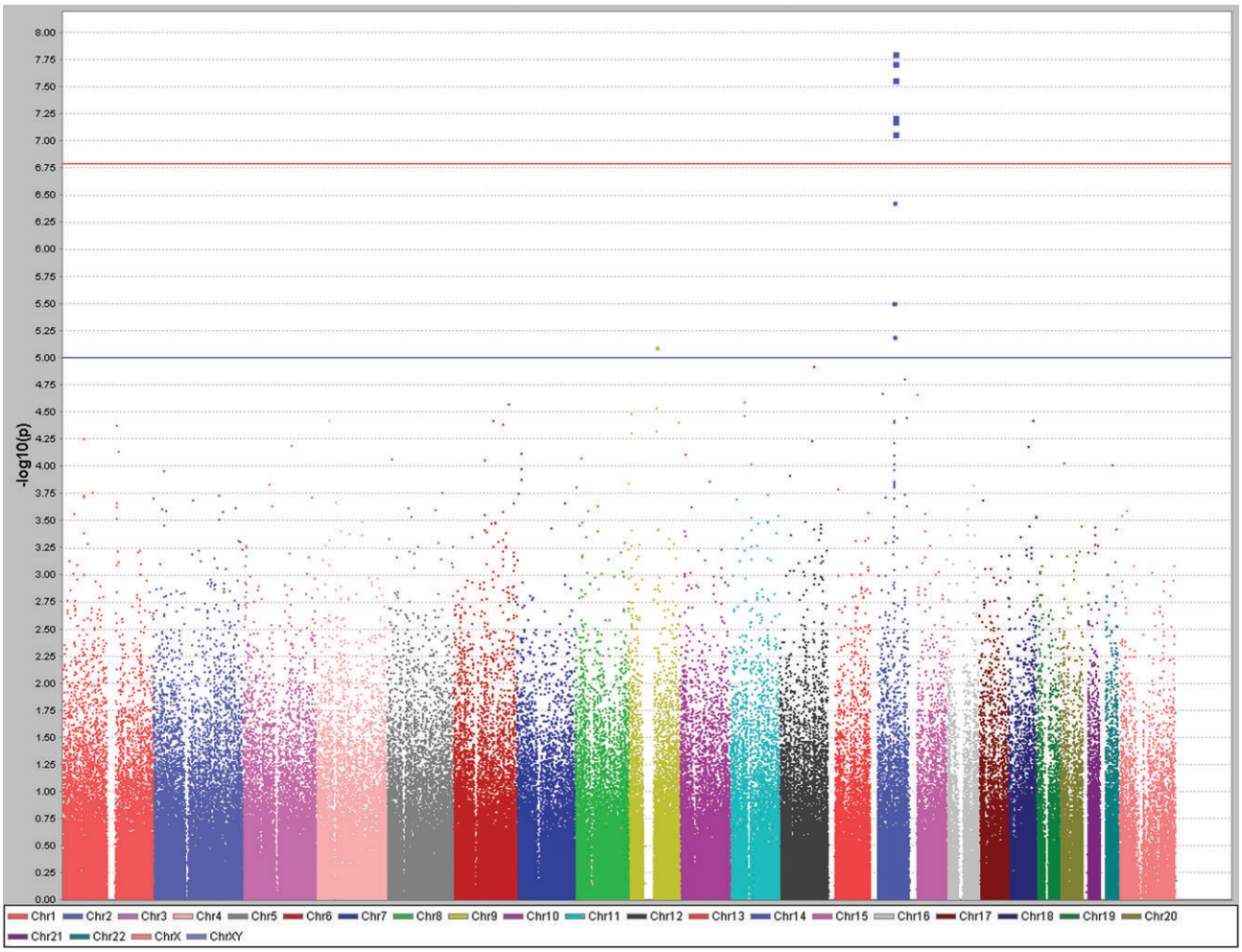

Fig. 3 Genome-wide association study (GWAS) results for N-linked A2 glycan (GlcNAc2Man3GlcNAc2) in human plasma. GWAS was performed on 924 people from the Croatian island of Vis.[53] The probability that each individual single nucleotide polymorphism (SNP) significantly affects level of A2 glycan is plotted for the whole genome (organized by individual chromosomes presented in different colours). There is a clear "peak" at chromosome 14 including 6 SNPs showing association with A2 glycan beyond genome-wide significance level after Bonferroni correction ($P = 1.7 \times 10^{-7}$) represented by the top horizontal line. The lower horizontal line represents a suggestive p-value of 10^{-5}.

high-throughput genomics and proteomics on the forefront of modern science—the "multi-dimensional biology". The analysis of both genetic polymorphisms and glycan levels in the same individuals will enable insights into complex systems of different transcription factors, transporters, enzymes, quality control chaperones and other proteins that regulate protein glycosylation. Determining the genetic polymorphisms in genes that code for these proteins will enable us to better understand molecular processes that underlay many diseases and provide a whole new set of targets for future pharmacological interventions.

Acknowledgements

The work in authors' laboratories is supported by grants #309-0061194-2023 (to GL) and #108-1080315-0302 (to IR) from the Croatian Ministry of Science, Education and Sport; by European Commission EUROPHARM, EUROSPAN, and INTEGERS grants the National Institute for Bioprocessing Research and Training, Ireland and Eurocarb DB. EuroCarbDB is a Research Infrastructure Design Study Funded by the 6th Research Framework Program of the European Union (Contract: RIDS Contract number 011952).

References

1 R. T. Lee, G. Lauc and Y. C. Lee, *EMBO Rep.*, 2005, **6**, 1018–1022.
2 R. G. Spiro, *Glycobiology*, 2002, **12**, 43R–56R.
3 M. Abu-Qarn, J. Eichler and N. Sharon, *Curr. Opin. Struct. Biol.*, 2008, **18**, 544–550.
4 A. Helenius and M. Aebi, *Science*, 2001, **291**, 2364–2369.
5 D. J. Moloney, L. H. Shair, F. M. Lu, J. Xia, R. Locke, K. L. Matta and R. S. Haltiwanger, *J. Biol. Chem.*, 2000, **275**, 9604–9611.
6 Y. C. Lee and R. T. Lee, *Acc. Chem. Res.*, 1995, **28**, 321–327.
7 G. Lauc and M. Heffer-Lauc, *Biochim. Biophys. Acta, Gen. Subj.*, 2006, **1760**, 584–602.
8 A. Varki, *Glycobiology*, 1993, **3**, 97–130.
9 R. S. Haltiwanger and J. B. Lowe, *Annu. Rev. Biochem.*, 2004, **73**, 491–537.
10 J. D. Marth and P. K. Grewal, *Nat. Rev. Immunol.*, 2008, **8**, 874–887.
11 N. Sharon, *J. Biol. Chem.*, 2007, **282**, 2753–2764.

12 R. Apweiler, H. Hermjakob and N. Sharon, *Biochim. Biophys. Acta, Gen. Subj.*, 1999, **1473**, 4–8.
13 G. Lauc, *Biochim. Biophys. Acta, Gen. Subj.*, 2006, **1760**, 525–526.
14 M. L. DeMarco and R. J. Woods, *Glycobiology*, 2008, **18**, 426–440.
15 K. W. Marek, I. K. Vijay and J. D. Marth, *Glycobiology*, 1999, **9**, 1263–1271.
16 H. H. Freeze, *Nat. Rev. Genet.*, 2006, **7**, 537–551.
17 J. Jaeken, *J. Inherited Metab. Dis.*, 2003, **26**, 99–118.
18 G. Vogt, A. Chapgier, K. Yang, N. Chuzhanova, J. Feinberg, C. Fieschi, S. Boisson-Dupuis, A. Alcais, O. Filipe-Santos, J. Bustamante, L. de Beaucoudrey, I. Al-Mohsen, S. Al-Hajjar, A. Al-Ghonaium, P. Adimi, M. Mirsaeidi, S. Khalilzadeh, S. Rosenzweig, O. de la Calle Martin, T. R. Bauer, J. M. Puck, H. D. Ochs, D. Furthner, C. Engelhorn, B. Belohradsky, D. Mansouri, S. M. Holland, R. D. Schreiber, L. Abel, D. N. Cooper, C. Soudais and J. L. Casanova, *Nat. Genet.*, 2005, **37**, 692–700.
19 G. Vogt, B. Vogt, N. Chuzhanova, K. Julenius, D. N. Cooper and J. L. Casanova, *Curr. Opin. Genet. Dev.*, 2007, **17**, 245–251.
20 R. A. Dwek, A. C. Lellouch and M. R. Wormald, *J. Anatom.*, 1995, **187**, 279–292.
21 R. B. Parekh, R. A. Dwek, B. J. Sutton, D. L. Fernandes, A. Leung, D. Stanworth, T. W. Rademacher, T. Mizuochi, T. Taniguchi and K. Matsuta, *Nature*, 1985, **316**, 452–457.
22 O. Gornik and G. Lauc, *Dis. Markers*, 2008, **25**, 267–278.
23 T. W. Rademacher, P. Williams and R. A. Dwek, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 6123–6127.
24 H. Albert, M. Collin, D. Dudziak, J. V. Ravetch and F. Nimmerjahn, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 15005–15009.
25 Y. Kaneko, F. Nimmerjahn and J. V. Ravetch, *Science*, 2006, **313**, 670–673.
26 M. Shibata-Koyama, S. Iida, A. Okazaki, K. Mori, K. Kitajima-Miyama, S. Saitou, S. Kakita, Y. Kanda, K. Shitara, K. Kato and M. Satoh, *Glycobiology*, 2009, **19**, 126–134.
27 A. Kobata, *Glycoconjugate J.*, 2000, **17**, 443–464.
28 Y. Hizukuri, Y. Yamanishi, K. Hashimoto and M. Kanehisa, *Genome Inform.*, 2004, **15**, 69–81.
29 L. Royle, E. Matthews, A. Corfield, M. Berry, P. M. Rudd, R. A. Dwek and S. D. Carrington, *Glycoconjugate J.*, 2008, **25**, 763–773.
30 O. Gornik, L. Royle, D. J. Harvey, C. M. Radcliffe, R. Saldova, R. A. Dwek, P. Rudd and G. Lauc, *Glycobiology*, 2007, **17**, 1321–1332.
31 L. Royle, M. P. Campbell, C. M. Radcliffe, D. M. White, D. J. Harvey, J. L. Abrahams, Y. G. Kim, G. W. Henry, N. A. Shadick, M. E. Weinblatt, D. M. Lee, P. M. Rudd and R. A. Dwek, *Anal. Biochem.*, 2008, **376**, 1–12.
32 A. Knezevic, O. Polasek, O. Gornik, I. Rudan, H. Campbell, C. Hayward, A. Wright, I. Kolcic, N. O'Donoghue, J. Bones, P. M. Rudd and G. Lauc, *J. Proteome Res.*, 2009, **8**, 694–701.
33 P. M. Rudd, I. Rudan and A. F. Wright, *J. Proteome Res.*, 2009, **8**, 1105–1105.
34 M. Flögel, G. Lauc, I. Gornik and B. Maček, *Clin. Chem. Lab. Med.*, 1998, **36**, 99–102.
35 I. Gornik, G. Maravić, J. Dumić, M. Flögel and G. Lauc, *Clin. Biochem.*, 1999, **32**, 605–608.
36 O. Gornik, I. Gornik, V. Gasparovic and G. Lauc, *Clin. Biochem.*, 2008, **41**, 504–510.
37 J. N. Arnold, R. Saldova, U. M. Hamid and P. M. Rudd, *Proteomics*, 2008, **8**, 3284–3293.
38 P. M. Visscher, W. G. Hill and N. R. Wray, *Nat. Rev. Genet.*, 2008, **9**, 255–266.
39 M. Abney, M. S. McPeek and C. Ober, *Am. J. Hum. Genet.*, 2000, **66**, 629–650.
40 C. Ober, M. Abney and M. S. McPeek, *Am. J. Hum. Genet.*, 2001, **69**, 1068–1079.
41 A. Wright, B. Charlesworth, I. Rudan, A. Carothers and H. Campbell, *Trends Genet.*, 2003, **19**, 97–106.
42 S. J. Chanock, T. Manolio, M. Boehnke, E. Boerwinkle, D. J. Hunter, G. Thomas, J. N. Hirschhorn, G. Abecasis, D. Altshuler, J. E. Bailey-Wilson, L. D. Brooks, L. R. Cardon, M. Daly, P. Donnelly, J. F. Fraumeni, Jr., N. B. Freimer, D. S. Gerhard, C. Gunter, A. E. Guttmacher, M. S. Guyer, E. L. Harris, J. Hoh, R. Hoover, C. A. Kong, K. R. Merikangas, C. C. Morton, L. J. Palmer, E. G. Phimister, J. P. Rice, J. Roberts, C. Rotimi, M. A. Tucker, K. J. Vogan, S. Wacholder, E. M. Wijsman, D. M. Winn and F. S. Collins, *Nature*, 2007, **447**, 655–660.
43 J. P. Ioannidis, *PLoS Med.*, 2005, **2**, e124.
44 L. R. Cardon and J. I. Bell, *Nat. Rev. Genet.*, 2001, **2**, 91–99.
45 V. Vitart, I. Rudan, C. Hayward, N. K. Gray, J. Floyd, C. N. Palmer, S. A. Knott, I. Kolcic, O. Polasek, J. Graessler, J. F. Wilson, A. Marinaki, P. L. Riches, X. Shu, B. Janicijevic, N. Smolej-Narancic, B. Gorgoni, J. Morgan, S. Campbell, Z. Biloglav, L. Barac-Lauc, M. Pericic, I. M. Klaric, L. Zgaga, T. Skaric-Juric, S. H. Wild, W. A. Richardson, P. Hohenstein, C. H. Kimber, A. Tenesa, L. A. Donnelly, L. D. Fairbanks, M. Aringer, P. M. McKeigue, S. H. Ralston, A. D. Morris, P. Rudan, N. D. Hastie, H. Campbell and A. F. Wright, *Nat. Genet.*, 2008, **40**, 437–442.
46 Y. S. Aulchenko, S. Ripatti, I. Lindqvist, D. Boomsma, I. M. Heid, P. P. Pramstaller, B. W. Penninx, A. C. Janssens, J. F. Wilson, T. Spector, N. G. Martin, N. L. Pedersen, K. O. Kyvik, J. Kaprio, A. Hofman, N. B. Freimer, M. R. Jarvelin, U. Gyllensten, H. Campbell, I. Rudan, A. Johansson, F. Marroni, C. Hayward, V. Vitart, I. Jonasson, C. Pattaro, A. Wright, N. Hastie, I. Pichler, A. A. Hicks, M. Falchi, G. Willemsen, J. J. Hottenga, E. J. de Geus, G. W. Montgomery, J. Whitfield, P. Magnusson, J. Saharinen, M. Perola, K. Silander, A. Isaacs, E. J. Sijbrands, A. G. Uitterlinden, J. C. Witteman, B. A. Oostra, P. Elliott, A. Ruokonen, C. Sabatti, C. Gieger, T. Meitinger, F. Kronenberg, A. Doring, H. E. Wichmann, J. H. Smit, M. I. McCarthy, C. M. van Duijn and L. Peltonen, *Nat. Genet.*, 2009, **41**, 47–55.
47 I. Rudan, A. Marusic, S. Jankovic, K. Rotim, M. Boban, G. Lauc, I. Grkovic, Z. Dogas, T. Zemunik, Z. Vatavuk, G. Bencic, D. Rudan, R. Mulic, V. Krzelj, J. Terzic, D. Stojanovic, D. Puntaric, E. Bilic, D. Ropac, A. Vorko-Jovic, A. Znaor, R. Stevanovic, Z. Biloglav and O. Polasek, *Croat. Med. J.*, 2009, **50**, 4–6.
48 J. N. Hirschhorn, *N. Engl. J. Med.*, 2009, **360**, 1699–1701.
49 D. B. Goldstein, *N. Engl. J. Med.*, 2009, **360**, 1696–1698.
50 H. Schachter, *Glycoconjugate J.*, 2000, **17**, 465–483.
51 P. Stanley, *Curr. Opin. Struct. Biol.*, 2007, **17**, 530–535.
52 A. V. Nairn, W. S. York, K. Harris, E. M. Hall, J. M. Pierce and K. W. Moremen, *J. Biol. Chem.*, 2008, **283**, 17298–17313.
53 G. Lauc, J. Huffman, C. Hayward, A. Knežević, O. Polašek, O. Gornik, V. Vitart, I. Kolčić, Z. Biloglav, L. Zgaga, N. D. Hastie, A. F. Wright, H. Campbell, P. M. Rudd and I. Rudan, *Nature Proceedings*, 2009, http://hdl.handle.net/10101/npre.2009.2864.1 .